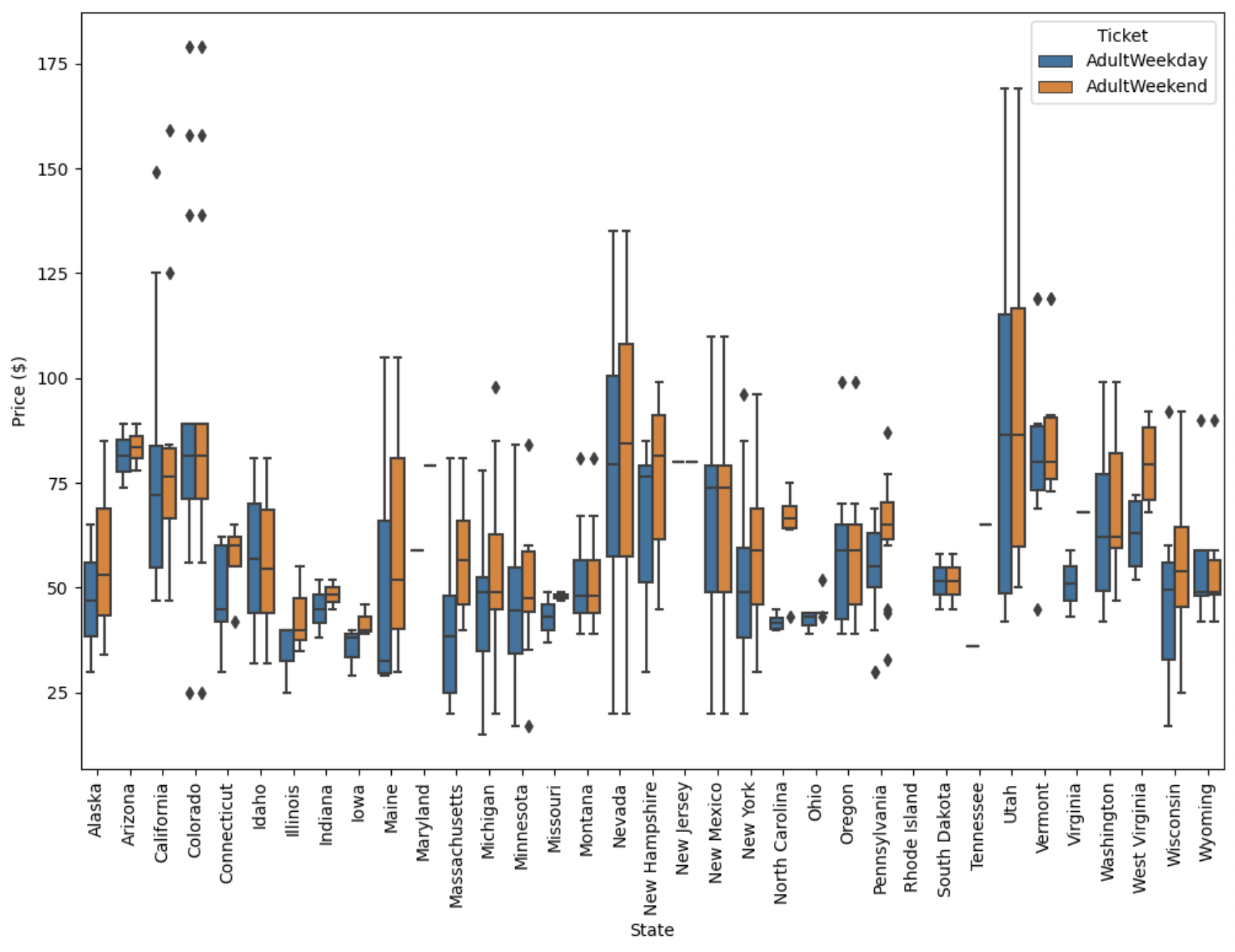
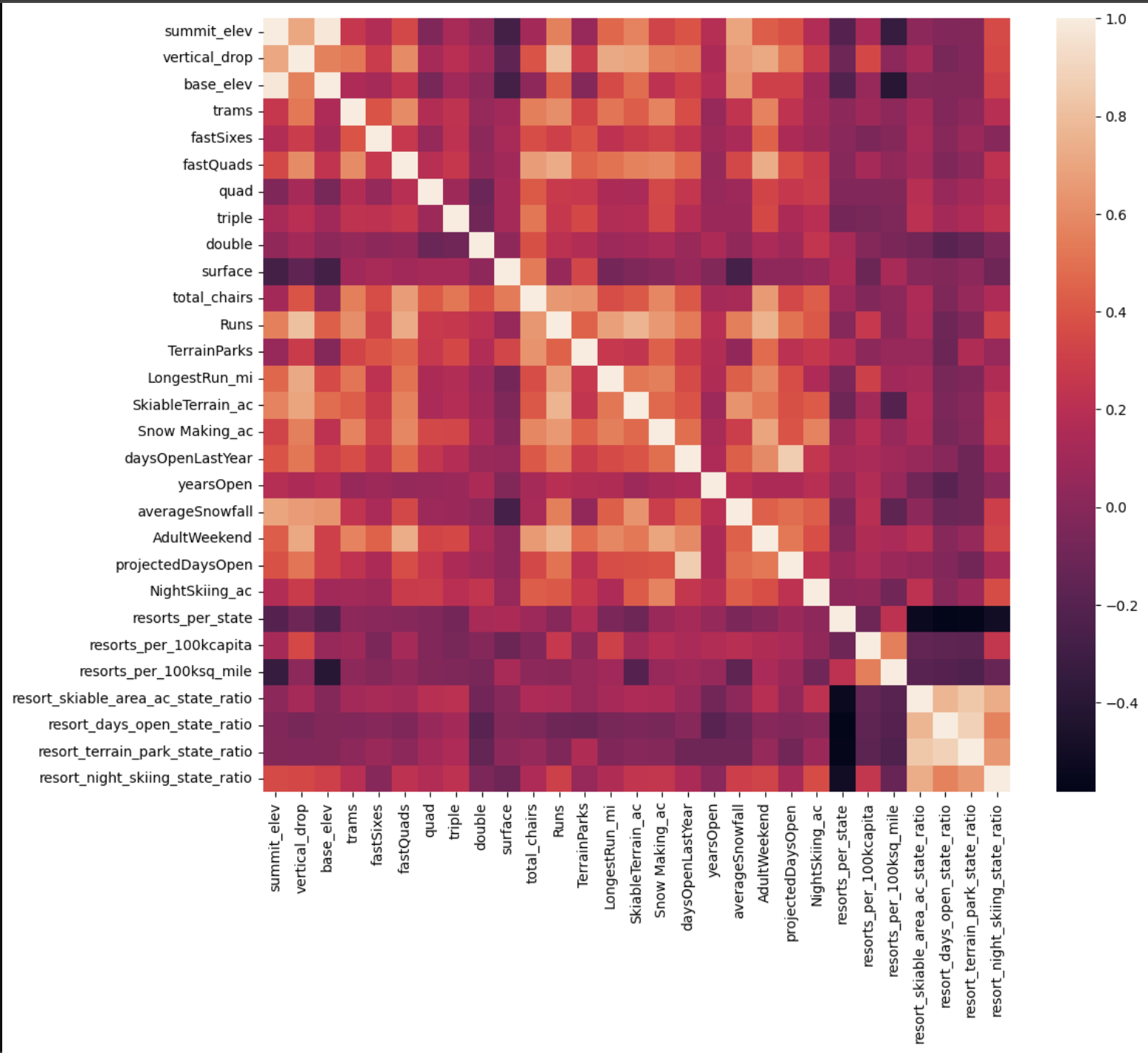
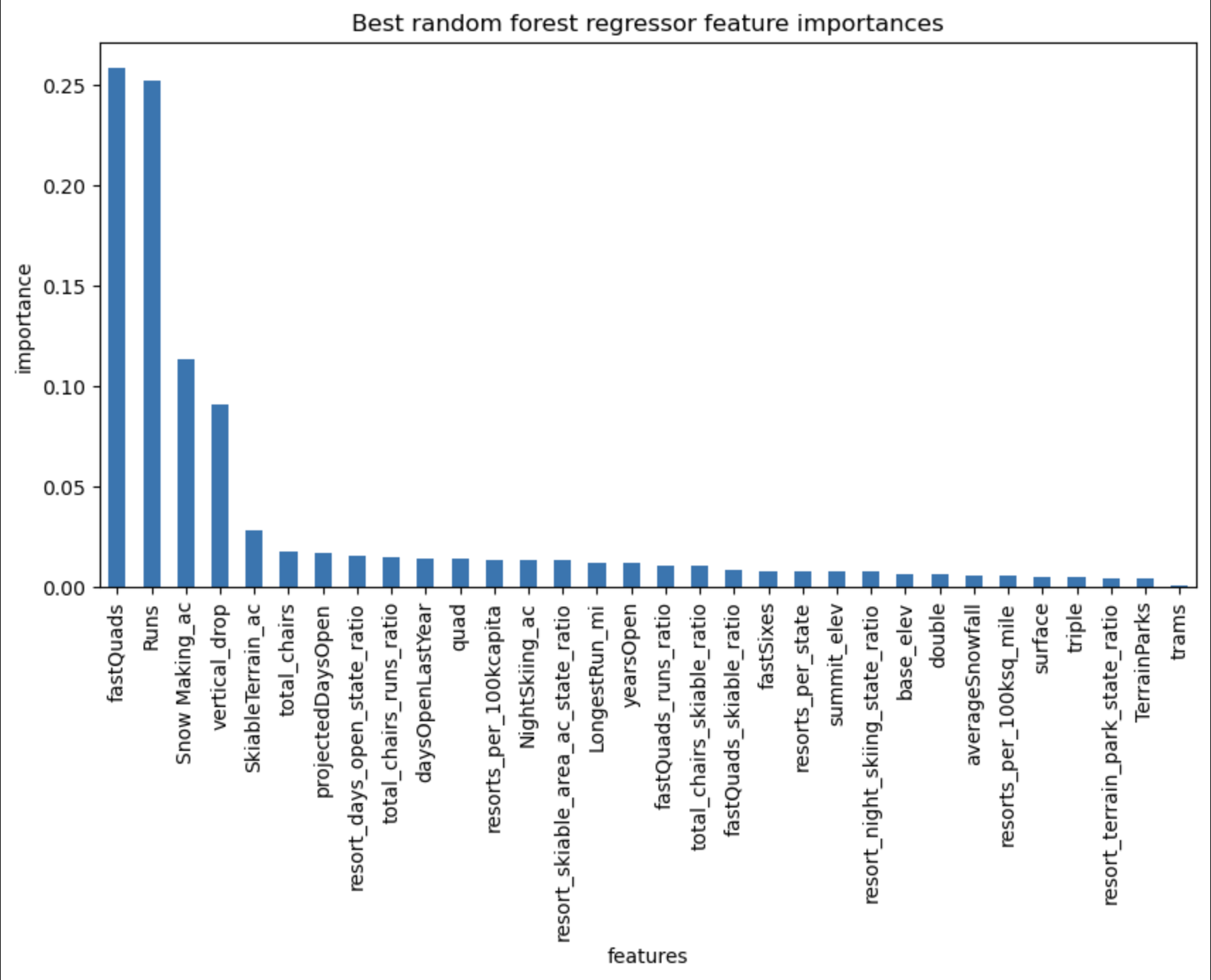
“Guided Capstone Project Report”

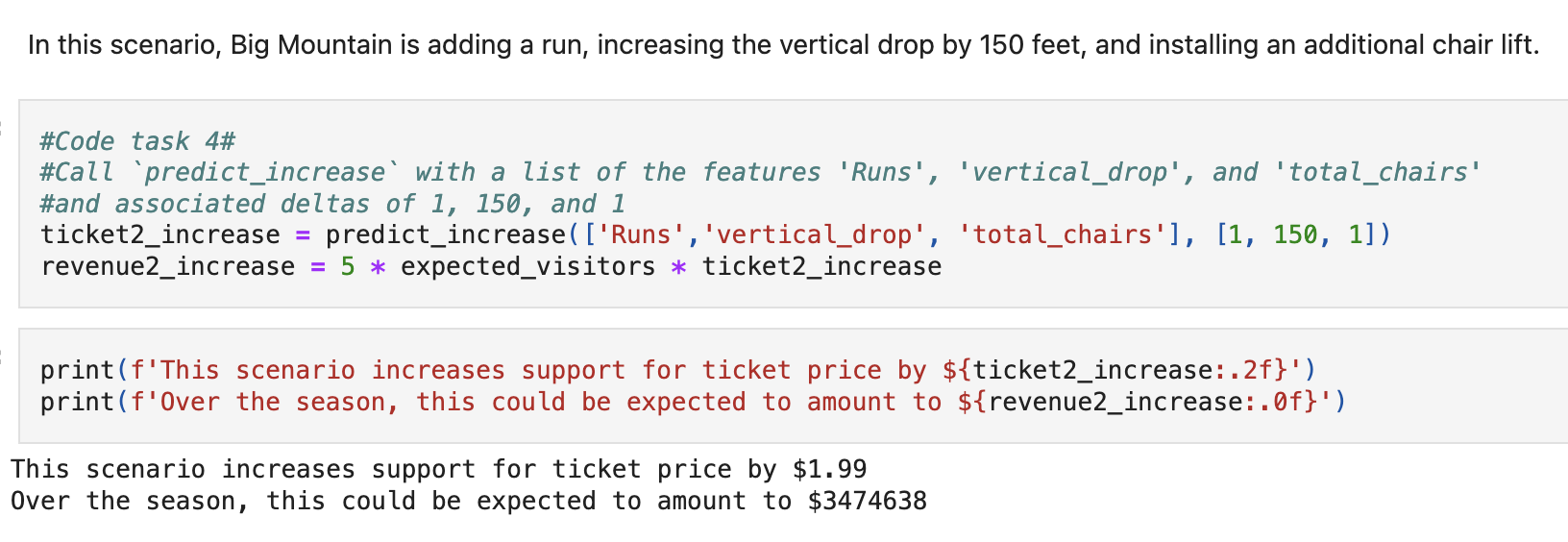
Big Mountain Resort is a ski resort that is looking to improve their finances. The resort suspects that there is potential to increase their revenue by increasing their ticket price. The best way to change the features of the resort in the most impactful way to increase revenue is to analyze the data of 330 ski resorts across the country. The role of the data scientist in this context is to identify the placement of the resort’s current ticket price versus the data generated ticket price, and the realistic parameters that the company can edit. Proposals for edits can look like changing the run amounts or the length of the runs, and also to predict how much the proposed changes can increase the ticket price based on the model constructed with the data. The data scientist has the responsibility to figure out the best plan of action.

In the wrangling of data, the ski data of 330 resorts were examined. The process of wrangling took place by examining the columns, the data types of columns, the missing values, percentage of missing values per column, and data duplicates. Following this examination, there was an attempt to see if the ticket price had a correlation with the state it was sold in. (Figure below is a bar plot of the ticket prices ber state) Throughout these preliminary visualizations of features per state, it was found that there were some exceptional data points, and then these outliers were removed in an effort to best represent the true spectrum of data points. At the end of the wrangling, there was an available data frame containing tidy and almost all numeric values ready for the next step.

The next step of the data science method is the exploratory data analysis, wherein the data frame was heavily visualized. Following the thorough investigation of the state versus ticket price correlation, it was found by component analysis that there was a lack of significant correlation between the two. Post this revelation, a heatmap was produced that showed the several true correlations found within the data frame in relation to the ticket price.A mass of scatterplots containing ticket price on the y axis and the other features of the data frame on the x axis revealed that the price had strong correlations with fast quads, runs, chairs, snow making, and night skiing. Given that there were clear frontrunners to where the resort could make some changes to increase their ticket price, it was time to generate the model to predict ticket prices.

The first step to model generation was the preprocessing of data. All data had to be numeric, all missing values had to be imputed by either the mean or median of the column, all data had to be scaled using the max, min, and mean, and finally the data had to be split between a train and test set. Once the data was processed, the regression was next. A linear regression pipeline produces an estimator, the linear regression, that can be used to generate predictions. The pipeline gets fitted by the training data set, and then the accuracy of the model is assessed by running the test results and evaluating the R squared value, mean absolute error value, and the standard deviation of the mean absolute error. The initial linear regression model was optimized by tuning its hyperparameters, but the tuning provided marginal increases in the model’s accuracy as assessed through the R squared value. So, a different estimator, the Random Forest Regressor, was used to generate a pipeline. With this new estimator, another round of hyperparameter tuning happened along with an analysis of which columns were most influential in the creation of the best predicting model. It was found that the Random Forest Regressor was able to provide a better R squared value and produce a smaller negative mean absolute error and negative mean absolute error standard deviation than the linear regression model. Additionally, the Random Forest Regressor provided the most influential columns in the formation of the model: fastQuads, Runs, SnowMaking\_ac, and vertical drop. Not only was the model created, but now there was even more confidence in specifically how the resort could influence their ticket price change.

Finally, the winning model could be used to generate the predicted price of a Big Mountain Resort ticket based on all of the data it had been trained on. But first, it needed one more thing to be trained on: the test data. Once the test data was uploaded to the training of the model, Big Mountain Resort was ready to be processed. The current ticket price of Big Mountain Resort is 81.00 dollars and the predicted price of the resort is 95.87 dollars. Not only was this predicted price generated, but a function was constructed to see how much a tweaking of certain features of Big Mountain Resort would impact price. With the model complete and the feature change to price impact function completed, it felt appropriate to make a recommendation on the ticket price change.

The final recommendation for the ticket price change was to increase from 81.00 dollars to 84.99 dollars along with changes in the physical landscape of the resort; installation of three more chairs along with increasing the vertical drop by 150 ft. (Below is a snippet of code demonstrating how the feature change to price impact function was used)Since the ticket price model predicted that Big Mountain Resort was undercharging, it was clear that the resort had potential to increase ticket prices right away. However, to ensure that the price change would accompany physical improvements to further justify the change, the feature change to price ticket change function was utilized. With the function, it was found that by adding 3 chairs and a 150 ft drop, that a price increase of 3.99 was permissible. With the operational costs of 4 new chairs total (around 6 million dollars), the increased revenue of 7 million dollars would cover the new expenses plus extra to spare.

The data used to generate the model and predict the ticket price has potential for including more data in the future; information that should be included is the mean/median income of the visitors, the distance the customers have to travel to get to the resort, and amenities provided for the customers. With information on aspects like these, there can be more definitive actions to improve the popularity of the resort and increase visitor number per year and frequency per visitor. New data on popularity practices in tandem with an increased ticket price can bolster Big Mountain Resort’s revenue greatly.

Using the data science method, an appropriate recommendation was generated on how to increase Big Mountain Resort’s revenue, along with guidance on how to further improve the data collection and revenue increase idea generation.